数理解析研究所講究録
270巻 1976年 8-13

万能 CF 文法 と 文法の準同型写像

京大数理解析研 笠井琢美

定義 1 $G=(N,\Sigma,P,S)$ を CF文法とする. ただし

N: *nonterminal symbols* の集合, Σ: *terminal symbols* の集合

P: *productions* の集合, S: *initial symbol*

$V=N\cup\Sigma$ とおく.

$$x_0 \overset{\rho_1}{\underset{\ell}{\Longrightarrow}} x_1 \overset{\rho_2}{\underset{\ell}{\Longrightarrow}} \cdots \overset{\rho_k}{\underset{\ell}{\Longrightarrow}} x_k$$

が G における *leftmost derivation* (x_{i-1} から x_i への変換に production ρ_i が適用される) で $\alpha=\rho_1\rho_2\cdots\rho_k\in P^*$ のとき

$$x_0 \overset{\alpha}{\underset{\ell}{\Longrightarrow}} x_k$$

と書く. さらに languages $L(G)\subset\Sigma^*$ と $A_\ell(G)\subset P^*$ を

$$L(G)=\{w\in\Sigma^* \mid S \overset{\alpha}{\underset{\ell}{\Longrightarrow}} w,\ \alpha\in P^*\}$$

$$A_\ell(G)=\{\alpha\in P^* \mid S \overset{\alpha}{\underset{\ell}{\Longrightarrow}} w,\ w\in\Sigma^*\}$$

で定義する.

注. $S \underset{\ell}{\overset{\alpha}{\Rightarrow}} w$, $w \in \Sigma^*$ のとき α を w の left parse といい $A_\ell(G)$ を G の left parse language という (Aho & Ullman [6]).
w の left parse α は w の derivation tree とみなすことができ, したがって $A_\ell(G)$ を G の derivation trees 全体からなる集合とみなすことができる.

定義 2　$G = (N, \Sigma, P, S)$ を CF 文法, $C \subset P^*$ とするとき

$$L_\ell(G; C) = \{ w \in \Sigma^* \mid S \underset{\ell}{\overset{\alpha}{\Rightarrow}} w, \ \alpha \in C$$

を C を control set として G によって生成される言語という.

定義 3　$G = (N, \Sigma, P, S)$ と $\bar{G} = (\bar{N}, \Sigma, \bar{P}, \bar{S})$ を CF文法とするとき G から $\bar{G}$ への left-homomorphism ($h : G \to \bar{G}$ と書く) とは monoid homomorphism $h : P^* \to \bar{P}^*$ で

$$h(A_\ell(G)) \subset A_\ell(\bar{G})$$

を満すものをいう.

注. 以後 left-homomorphism を単に homomorphism と呼び, A_ℓ, $\underset{\ell}{\Rightarrow}$, L_ℓ などの添字 ℓ は省略する.

homomorphism $h : G \to \bar{G}$ が proper であるとは任意の $w \in \Sigma^*$ に対し

$$S \overset{\alpha}{\Longrightarrow} w \quad \text{in } G \quad \text{iff} \quad \bar{S} \overset{h(\alpha)}{\Longrightarrow} w \quad \text{in } \bar{G}$$

が成立するときをいう.

$h : G \to \bar{G}$ が proper な homomorphism なら

$$L(\bar{G} : h(A(G))) = \{ w \in \Sigma^* \mid \bar{S} \overset{h(\alpha)}{\Longrightarrow} w \text{ in } \bar{G},\ \alpha \in P^* \}$$
$$= L(\bar{G} : h(P)^*)$$

となることに注意する.

目的　以下で次のようなCF文法 G_U の存在を示す: 任意のCF文法 G に対し proper な homomorphism $h : G \to G_U$ が存在する. この結果の一般の文法への拡張は Hart [3] によって試みられている.

Notation　Σ を与えられた alphabet とする. (以後 Σ は固定する.) $c, \bar{c}, d, \bar{d}$ を Σ の元でない symbol とし

$$\Delta = \Sigma \cup \{c, \bar{c}, d, \bar{d}\}$$

とおく. $\sim$ を次で定義される Δ^* 上の関係とする.

$$x c \bar{d} \bar{c} y \sim x \bar{d} y \qquad x, y \in \Delta^*$$
$$x d \bar{d} \bar{d} y \sim x \bar{d} y \qquad x, y \in \Delta^*$$
$$x a y \sim x y \qquad a \in \Sigma,\ y \in (\Delta - \bar{c})\Delta^*$$

$\overset{*}{\sim}$ を $\sim$ の reflexive transitive closure とし

$$D = \{ w \in \Delta^* \mid w \overset{*}{\sim} \bar{d} \}$$

とおく.

<u>定義 4</u>　$G = (\{X_0, X_1, \cdots, X_n\}, \Sigma, P, X_0)$ を任意の CF文法とする.

$$\xi : V^* \to \Delta^* \qquad h : P^* \to \Delta^*$$

を次のように定義する.

$$\begin{cases} \xi(a) = a & a \in \Sigma \\ \xi(X_i) = d c^i & 0 \le i \le n \end{cases}$$

$$h(X_i \to x) = \bar{c}^i \, \xi(x^R) \, \bar{d}, \qquad X_i \to x \in P.$$

<u>Lemma 1.</u>　$X_i \overset{\alpha}{\Longrightarrow} w$, for some $w \in \Sigma^*$

iff　$d c^i \bar{d} \, h(\alpha) \overset{*}{\sim} \bar{d}$

<u>系</u>　$\alpha \in A(G)$　iff　$h(\alpha) \in D$

<u>定義 5</u>　$G_U = (\{S, X\}, \Sigma, P_U, S)$

$$P_U = \{ a : S \to Sa \mid a \in \Sigma \}$$
$$\cup \left\{ \begin{array}{ll} c : S \to SX, & \bar{c} : X \to \lambda \\ d : S \to SS, & \bar{d} : S \to \lambda \end{array} \right\}$$

以後 P_U と Δ を同一視する. すなわち production $\rho : X \to x \in P_U$ を文字 $\rho \in \Delta$ とみなす.

<u>Lemma 2</u> $A(G_U) = D$

<u>系</u> $h: P^* \to \Delta^*$ ($\Delta = P_U$) は G から G_U への homomorphism.

<u>Theorem</u> $h: G \to G_U$ は proper homomorphism.

<u>系1</u> 任意の context-free languag $L \subset \Sigma^*$ に対し,

$$L = L(G_U; C)$$

となる regular control set C が存在する. (特に C を $C = \{\alpha_1, \alpha_2, \cdots, \alpha_\ell\}^*$, $\alpha_i \in P_U^*$, に選ぶことができる. matrix grammar)

<u>系2</u> $\{ L(G_U ; C) \mid C$ は regular control set $\}$

= Σ 上の context-free languages の族.

<u>系3</u> 任意の CF文法 G と G の regular control set C_1, C_2 に対し $L(G : C_1) = L(G : C_2)$?, $L(G : C_1) \cap L(G : C_2) = \phi$?, $L(G : C_1) - L(G ; C_2)$ は finite ? などは recursively unsolvable.

注 G が unambiguous なら 系3 の問題は solvable となる.

Acknowledgment:

有益な御討論をしていただきました、高須達教授と林健志氏に感謝いたします。

文献

[1] T. KASAI A hierarchy between context-free and context-sensitive languages JCSS 4 (1970

[2] T. KASAI A universal context-free grammar Information & Control 28 (1975)

[3] J.M. Hart Two extension of KASAI's universal context-free grammar. Univ. of Kentucky. technical report No 22-75

[4] J.M Hart Derivation languages and syntactical categories. Information & Control 28 (1975)

[5] R.L. Cannon State grammar sarsing Doctral Dissertation. Univ. of North Carolina.

[6] A.V. Aho and J.D. Ullman The Thory of Parsing Translation and Compiling.